

Coursera Capstone Project

IBM Applied Data Science Capstone

Pratik Shukla

21-09-2020

Description of the data and how it will be used to solve the problem.

2. Data Section:

Description of the data and its sources that will be used to solve the problem

Description of the Data:

The following data is required to answer the issues of the problem:

- List of Boroughs and neighborhoods of Manhattan with their geodata (latitude and longitude)
- List of Subway metro stations in Manhattan with their address location
- List of apartments for rent in Manhattan area with their addresses and price
- Preferably, a list of apartment for rent with additional information, such as price, address, area, # of beds, etc
- Venues for each Manhattan neighborhood (than can be clustered)
- Venues for subway metro stations, as needed

How the data will be used to solve the problem

The data will be used as follows

:

- Use Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups (as per Course LAB)
- Use foursquare and geopy data to map the location of subway metro stations , separately and on top of the above clustered map in order to be able to identify the venues and amenities near each metro station, or explore each subway location separately
- Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations.
- create a map that depicts, for instance, the average rental price per square ft, around a radius of 1.0 mile (1.6 km) around each subway station - or a similar metric. I will be able to quickly point to the popups to know the relative price per subway area.
- Addresses from rental locations will be converted to geodata(lat, long) using Geopy-distance and Nominatim.
- Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies such as Metro New York MTA, etc

The processing of these DATA will allow to answer the key questions to make a decision:

- What is the cost of rent (per square ft) around a mile radius from each subway metro station?
- What is the area of Manhattan with best rental pricing that meets criteria established?
- What is the distance from the workplace (Park Ave and 53 rd St) and the tentative future home?
- What are the venues of the two best places to live? How do the prices compare?
- How are venues distributed among Manhattan neighborhoods and around metro stations?
- Are there tradeoffs between size and price and location?
- Any other interesting statistical data findings of the real estate and overall data.

2.1 Data Requirements

- Geodata for current residence in Singapore with venues established using Foursquare.
- List of Manhattan (MH) neighborhoods with clustered venues established via Foursquare (as in CourseLab).
https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods#Midtown_neighborhoods
- List of subway metro stations in Manhattan with addresses and geo data (lat,long):
https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan ,
(<https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1>)
- List of apartments for rent in the Manhattan area with information on neighborhood location, address,number of beds, area size, monthly rent price and complemented with geo data via Nominatim.
<http://www.rentmanhattan.com/index.cfm?page=search&state=results>
<https://www.nestpick.com/search?city=new-york>
- Place to work in Manhattan (Park Avenue and 53rd St) for reference

